

---

## Application Note

# SubNet: a Java application for subnetwork extraction

Quanwei Zhang, Zhengdong D. Zhang

Department of Genetics, Albert Einstein College of Medicine, 1301 Morris Park Avenue, Bronx, NY 10461, USA

---

## SUPPLEMENTARY DATA

**Supplementary Table 1.** Performance comparison among the four methods.

Network type	Method	Seeds	Time	Memory	Network type	Method	Seeds	Time	Memory
Undirected	Shell-1	2	5.10 s	735 MB	Directed	Shell-1	2	4.77 s	742 MB
		5	6.11 s	743 MB			5	5.26 s	742 MB
		10	6.32 s	751 MB			10	5.81 s	742 MB
		50	10.89 s	833 MB			50	8.56 s	744 MB
		100	15.72 s	834 MB			100	11.75 s	839 MB
	Shell-2	2	6.28 s	741 MB		Shell-2	2	4.97 s	741 MB
		5	6.93 s	744 MB			5	5.23 s	745 MB
		10	9.19 s	752 MB			10	6.78 s	751 MB
		50	22.70 s	1.3 GB			50	13.05 s	842 MB
		100	28.67 s	1.4 GB			100	16.79 s	1.3 GB
	Shortest path	2	6.34 s	770 MB		Shortest path	2	2.84 s	239 MB
		5	8.75 s	776 MB			5	7.47 s	464 MB
		10	28.45 s	1.4 GB			10	28.7 s	673 MB
		50	6.62 m	3.6 GB			50	7.77 m	2.4 GB
		100	26.97 m	4.5 GB			100	27.83 m	4.3 GB
	Emission decay (linear)	2	6.59 h	2.4 GB		Emission decay (linear)	2	1 h 34.43 m	2.3 GB
		5	17.54 h	2.5 GB			5	3.36 h	2.4 GB
		10	1 d 9.49 h	4.3 GB			10	7.37 h	3.6 GB
		50	7 d 2 h	4.3 GB			50	1 d 12.9 h	4.3 GB
		100	14 d 4.05 h	4.5 GB			100	2 d 22.63 h	4.5 GB
PageRank	2	28.93 s	5.6 GB	PageRank	2	28.84 s	5.6 GB		
	5	28.05 s	5.6 GB		5	28.94 s	5.6 GB		
	10	28.28 s	5.6 GB		10	29.17 s	5.6 GB		
	50	28.49 s	5.6 GB		50	29.1 s	5.6 GB		
	100	32.27 s	5.6 GB		100	32.76 s	5.6 GB		

**Notes:**

1. The benchmark used for the comparison was the HPRD protein interaction network (version 9, 9,671 nodes and 37,058 edges). Five sets of 2, 5, 10, 50, and 100 randomly selected genes ('seeds') were first generated and then used for all comparisons.
2. The benchmarking was performed on a Linux Dell PowerEdge 815 server with 4 2200-MHz AMD Opteron CPUs and 256GB memory.
3. The comparison shows that the shell and the shortest path perform relatively fast with a small memory footprint despite the large size of the network. The PageRank method requires more memory but performs fast. The intensive calculation needed by the decay methods makes it more time and memory consuming.

**Supplementary Table 2.** Comparison of the four subnetwork extraction methods implemented in SubNet.

	Shell	Shortest path	Emission decay	PageRank
Network topology used	Local	Global	Local and global	Local and global
Can edge weights be used?	No	Yes	Yes	Yes
Computational intensity	Not intense	Intense	Intense	Intense
Speed	Fast	Slow	Slow	Fast
Memory usage	Low	Medium low	Medium high	High
Scalability with number of seeds	Very scalable	Not very scalable	Not very scalable	Very scalable

## Notes:

1. Scalability measures how the method performs (e.g., computational intensity) as the number of seeds increases.
2. Broadly speaking, there are two classes of computational methods for subnetwork identification. One class identifies subnetworks (a.k.a. modules) by using the internal topology (wiring) of the network alone. There are several ways to achieve this, e.g., by simulated annealing (Ideker et al, 2002, *Bioinformatics*) or spectral modularity maximization (Newman, 2006, *PNAS*). The other class identifies subnetworks based on both a list of selected genes and the network. The methods implemented in SubNet and reviewed in the paper belong to this class.

**Supplementary Table 3.** Top 100 nodes of a subnetwork extracted with the PageRank method.

Rank	Protein symbol	PageRank score	Degree in subnetwork	Rank	Protein symbol	PageRank score	Degree in subnetwork
1	HMGB1	0.03191223	21	51	HDAC1	9.52E-04	14
2	XRCC4	0.009704615	4	52	CRMP1	9.51E-04	2
3	LIG4	0.008160946	2	53	PLG	9.37E-04	1
4	HMGB2	0.005228549	5	54	ATN1	9.35E-04	8
5	LIG1	0.003037483	2	55	POU5F1	9.33E-04	4
6	TP53	0.00283791	26	56	VIM	9.32E-04	11
7	YWHAG	0.002730851	13	57	LIG3	9.27E-04	0
8	ATXN1	0.002294739	7	58	PTN	9.24E-04	2
9	UBQLN4	0.002098236	8	59	CDK1	9.11E-04	13
10	PRKCA	0.00191381	16	60	RAG1	8.97E-04	2
11	CSNK2A1	0.001767497	13	61	SUMO4	8.95E-04	1
12	EP300	0.001640596	21	62	MAGEA11	8.91E-04	2
13	TGFBR1	0.001630948	10	63	C14orf1	8.85E-04	2
14	EWSR1	0.001628663	13	64	PTPN11	8.85E-04	11
15	AR	0.001608212	25	65	MAPK3	8.81E-04	11
16	SRC	0.001605253	23	66	SHC1	8.73E-04	16
17	CREBBP	0.00159708	17	67	TRAF6	8.69E-04	5
18	SMAD3	0.001585346	24	68	TLE1	8.68E-04	3
19	GRB2	0.001569997	12	69	PTPN6	8.64E-04	14
20	SMAD9	0.001546658	7	70	UBE2I	8.60E-04	12
21	CCDC85B	0.001509808	3	71	APP	8.57E-04	5
22	SMAD2	0.001502571	20	72	PRKCD	8.56E-04	24
23	CASP3	0.001490922	13	73	PPP1CA	8.52E-04	7
24	RB1	0.00147026	16	74	GF1B	8.40E-04	3
25	SMAD4	0.001465772	22	75	KRTAP4-12	8.32E-04	4
26	ESR1	0.001448517	29	76	JUN	8.29E-04	17
27	TRAF2	0.001431625	5	77	PLSCR1	8.29E-04	8
28	EGFR	0.001410965	20	78	SMURF1	8.26E-04	3
29	PRKACA	0.001405785	9	79	SKIL	8.24E-04	8
30	MDFI	0.001339353	3	80	EEF1A1	8.21E-04	11
31	CTNNB1	0.001301441	19	81	ABL1	8.19E-04	20
32	C1orf103	0.001243517	2	82	RAC1	8.18E-04	8
33	RELA	0.001234059	14	83	CEBPB	8.17E-04	13
34	FYN	0.001206961	16	84	SHBG	8.12E-04	2
35	MAPK1	0.001201822	17	85	LYN	7.89E-04	13
36	YWHAB	0.001166774	6	86	BRCA1	7.89E-04	20
37	UNC119	0.001127094	7	87	RAF1	7.84E-04	16
38	ACTB	0.001109013	11	88	HNRNPK	7.82E-04	9
39	CALM1	0.001104184	7	89	FN1	7.81E-04	4
40	SMAD1	0.001103343	10	90	ATF7IP	7.81E-04	3
41	TBP	0.001087493	7	91	HSP90AA1	7.72E-04	16
42	CSNK1A1	0.001083656	7	92	SETDB1	7.68E-04	5
43	PRKDC	0.001082133	14	93	LCK	7.68E-04	15
44	YWHAZ	0.001062425	10	94	NDRG1	7.63E-04	4
45	PCNA	0.001054191	6	95	PLCG1	7.56E-04	11
46	NFKB1	9.93E-04	9	96	ZBTB16	7.52E-04	12
47	AKT1	9.85E-04	16	97	DYNLL1	7.50E-04	5
48	DLG4	9.84E-04	6	98	PRPF40A	7.43E-04	6
49	ACTA1	9.77E-04	6	99	TP73	7.37E-04	8
50	PIK3R1	9.77E-04	17	100	SMURF2	7.36E-04	7

## Notes:

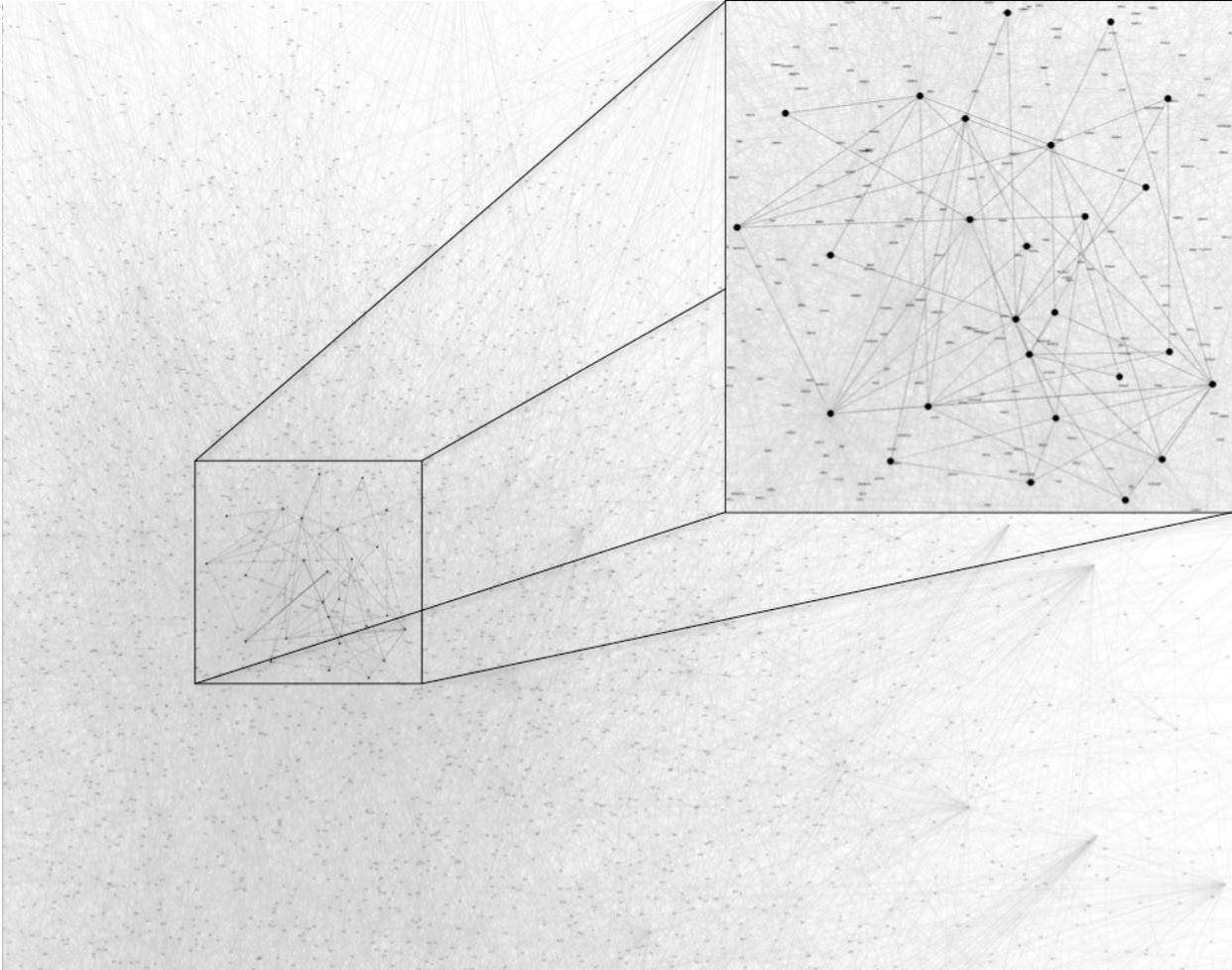
1. The basal network was HPRD protein interaction network (version 9).
2. The seeds are six DNA ligation-related proteins: HMGB1, HMGB2, LIG1, LIG3, LIG4, and XRCC4.
3. The PageRank method was run with parameters:  $c = 100$  (default) and  $d = 0.85$ .

**Supplementary Table 4.** Nodes of a subnetwork extracted with the shortest path method.

Protein symbol	Degree in subnetwork	Rank in Supplementary Table 3
NFKB1	4	46
PCNA	6	45
CSNK2A1	8	11
TP53	7	6
POU5F1	2	55
GZMB	3	
LIG4	3	3
PARP1	10	
APEX1	5	
POU2F1	4	
RELA	7	33
RAG1	2	60
LIG1	2	5
HMGB1	9	1
CSNK1A1	3	42
XRCC4	7	2
PNKP	2	
LIG3	2	57
IRF2	4	
HMGB2	8	4
APLF	3	
XRCC1	8	
DNMT1	2	
PGR	4	
APTX	3	
PRKDC	8	43

## Notes:

1. The basal network was HPRD protein interaction network (version 9).
2. The seeds are six DNA ligation-related proteins: HMGB1, HMGB2, LIG1, LIG3, LIG4, and XRCC4, the same as in Supplementary Table 3.



**Supplementary Figure 1.** Subnetwork extraction by SubNet. A subnetwork, highlighted with black nodes and edges in the zoom-in window, was extracted based on six DNA ligation-related proteins from the basal HPRD protein-protein interaction network (in grey).